DOCUMENT RESUME

ED 183 598 TM 010 119

AUTHOR	Price, Gary G.: And Others
TITLE	Procedures Used in Phase I of the IGE Evaluation to Scale, Adjust, and Aggregate Reading and Math Achievement in Grade 2. Report from the Project on Evaluation of Practices in Individualized Schooling.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCIS-TR-508
PUB DATE	Nov 79
GRANT	OB-NIE-G-78-0217
NOTE	40p.: For related documents, see TM 010 047-060 and TM 010 118.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Ability: *Achievement Gains: Achievement Tests: Elementary School Mathematics: Grade 2: Individualized Instruction: Internal Scaling: Multiple Regression Analysis: Primary Education: Reading Achievement: *Research Methodology: Scores: *Student Testing: *Test Interpretation
IDENTIFIERS	Comprehensive Tests of Basic Skills: Evaluation of Practices Individualized Schooling: Individually Guided Education: Short Form Test of Academic Aptitude: *Test Analysis: Winsorization

ABSTRACT

The procedures used in Phase I of the Individually Guided Education (IGE) evaluation to scale, adjust, and aggregate reading and mathematics achievement in grade 2 are described. An explanation is included of the administration procedures. The tests used (the Comprehensive Tests of Basic Skills and the Short Form Test of Academic Aptitude) and the adjustment made on individual student scores for academic ability and background variables are discussed. An explanation of how the achievement scores of students within a school were aggregated, and a description of the data are also included. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183598

IM010 119

TM

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Technical Report No. 508

Procedures Used in Phase 1 of the IGE Evaluation to Scale, Adjust, and Aggregate Reading and Math Achievement in Grade 2

by Gary G. Price, Herbert Van Deventer, Terence C. Janicki, and Thomas A. Romberg

November 1979

Wisconsin Research and Development Center for Individualized Schooling

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Rossmiller
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Technical Report No. 508

PROCEDURES USED IN PHASE I OF THE IGE EVALUATION TO SCALE, ADJUST, AND AGGREGATE READING AND MATH ACHIEVEMENT IN GRADE 2

by

Gary G. Price, Herbert C. Van Deventer, Jr.
Terence C. Janicki, and Thomas A. Romberg

Report from the Project on
Evaluation of Practices in Individualized Schooling

Gary G. Price and Thomas A. Romberg
Faculty Associates

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

November 1979

Published by the Wisconsin Research and Development Center for Individualized Schooling. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Grant No. OB-NIE-G-78-0217

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

Table of Contents

	Page
List of Tables	vii
List of Figures	ix
Abstract	xi
Tests Given to Pupils	2
A Problem with the Mathematics Computation Subtest	4
Adjustment of Student Achievement Scores	5
Guarding Against Faults in SFTAA Scores	9
Winsorization	9
Achievement Scores as Regression Residuals	15
Aggregation of Regression Residuals	17
References	27

List of Tables

Table		Page
1	Grade 2 Student Testing Program and Schedule	3
2	Grade 2 Multiple Regression Coefficients	16
3	Grade 2 Student Ability and Achievement.	25
4	Grade 2 Student Ability and Achievement Descriptive Statistics for School Aggregates . . .	26

List of Figures

Figure		Page
1	SFTAA bivariate distribution before winsorization. .	12
2	SFTAA bivariate distribution after winsorization . .	13
3	Reading vocabulary (CTBS) grade 2.	18
4	Reading comprehension (CTBS) grade 2	19
5	Mathematics computation (CTBS) grade 2	20
6	Mathematics concepts (CTBS) grade 2.	21
7	Mathematics applications (CTBS) grade 2.	22
8	Adjusted reading achievement grade 2	23
9	Adjusted mathematics achievement grade 2	24

Abstract

This technical report describes the procedures used in Phase I of the Individually Guided Education (IGE) evaluation to scale, adjust, and aggregate reading and mathematics achievement in grade 2. It includes an explanation of the administration procedures, a list of the tests administered to pupils, and a description of the adjustments made on individual student scores. A discussion of how the achievement scores or students within a school were aggregated, and a description of the data are also included.

This technical report describes the procedures used in Phase I of the Individually Guided Education (IGE) evaluation to scale, adjust, and aggregate reading and mathematics achievement in grade 2. A similar description of the procedures used with grade 5 data is given in Technical Report No. 509. An overview of the Phase I study and its relation to other phases of the IGE evaluation is provided by Romberg (1976). For an explanation of how schools were selected to participate in the Phase I study, see Price (1977). The present report includes an explanation of the administration procedures, a list of the tests administered to pupils, and a description of the adjustments made on individual student scores. A discussion of how the achievement scores of students within a school were aggregated, and a description of the data are also included.

Achievement scores, of course, do not tell the whole story of a school program's success. Even if students achieve high test scores, the value of that program might be questioned if teachers dislike its approach, or if students exhibit socially undesirable behavior, lack of creativity, or a negative attitude toward subject matter. Consequently, the Phase I study considers other outcomes, such as teachers' job satisfaction, students' self-concept, and students' self-direction. This document, however, is confined to measures of reading and mathematics achievement.

Tests Given to Pupils

A number of aptitude and achievement tests were administered at each school in October 1977. All participating students (N = 5,352) took the Short Form Test of Academic Aptitude (SFTAA), Level 1. The SFTAA yielded a general academic ability score and subtest scores for language and nonlanguage abilities. The Comprehensive Tests of Basic Skills (CTBS), Form S, Level C, Expanded Edition, provided measures of reading and mathematics achievement. The following subtests of the CTBS were administered: reading vocabulary, reading comprehension of sentences, reading comprehension of passages, mathematics computation, and mathematics concepts and applications. Because of time constraints, most students took only some of the CTBS subtests. Two personality measures, self-observation scales and Bialer's (1961) locus of control scale, were also administered and will be discussed in a later report. A list of the tests administered to grade 2 students is included in Table 1. The rationale for selecting these tests has been described by Stewart (1977).

It was unnecessary to require each student to take every test, since the grade 2 instruction and research (I&R) unit, rather than the individual student, was the unit of analysis. Tests were assigned to four groups of students, so that testing time required for any one group would not exceed 2 hours. The entire SFTAA was taken by all four groups, whereas the five subtests of the CTBS, the self-observation scales, and Bialer's locus of control scale were divided among the groups. The actual time requirement for grade 2 students was approximately 90 minutes. The tests taken by each group were spread over three sessions, as shown in Table 1.

Table 1

Grade 2 Student Testing Program and Schedule

Test session	Test group	Test	Working time for tests that have time limits (in minutes)	Approximate total time (in minutes)[a]
1	A	SFTAA[b]--Tests 1, 2, 3, 4	28	31
	B	SFTAA--Tests 1, 2, 3, 4	28	31
	C	SFTAA--Tests 1, 2, 3, 4	28	31
	D	SFTAA--Tests 1, 2, 3, 4	28	31
2	A	CTBS[c] Test 1--Reading vocabulary		20
	B	CTBS Test 3--Reading comprehension of passages	21	26
	C	CTBS Test 8--Mathematics concepts and applications		30
	D	CTBS Test 7--Mathematics computation	34	49
3	A	CTBS Test 2--Reading comprehension of sentences	20	25
	B	Self-observation scales	No limit	25
	C	Bailer's locus of control scale	No limit	25
	D	Test group D took no test in test session 3	--	--

[a]The approximate total time for each test session includes only average instruction and working time. It does not include time required for recording identifying data on the test books (completed by teachers or aides), assembling students into test groups, and distributing and collecting test materials.

[b]SFTAA--Short Form Test of Academic Aptitude, Level 1.

[c]CTBS--Comprehensive Tests of Basic Skills, Form S, Level C.

One grade 2 I&R unit per school was included in the study. A portion of the grade 2 children in each of those I&R units was then selected for testing. The selected students were randomly assigned to four test groups of approximately equal size. No more than 50 students in an I&R unit were tested; consequently, no test group had more than 13 students. The mean was approximately 9 students per test group. To assure a reasonably precise estimate of a unit's mean on a subtest, a minimum of 6 students per test group was required. In units with fewer than 24 students, this requirement made it necessary that some students take more tests than usual. Of the 156 grade 2 units, 22 had fewer than 24 students. For 15 of those units, tests were divided among three test groups, rather than four. For five units, tests were divided among two test groups. In two especially small units, each child took every test. Procedures used for special cases are described by Klopp, Buchanan, Stewart, and Romberg (1979).

A Problem with the Mathematics Computation Subtest

Staff members of six schools decided not to give their grade 2 students an eight-item multiplication section of the CTBS mathematics computation subtest. Because these second grade students had not had an opportunity to do all items, it would have been misleading to compare their raw scores with those of children from other schools. To make scores from the six schools comparable with scores from the other 150 schools, we could have omitted the multiplication section from all schools. But too much useful information would have been discarded with that approach. Instead, we computed a regression equation, by predicting the number of items correct in the multiplication section from the number of items correct on the

remainder of the mathematics Computation subtest. This regression equation used data from 1,341 children in schools giving the entire subtest. The regression equation, which has an R^2 of .257 is:

$$\hat{Y} = X + (.288X + .527).$$

$\hat{Y}$ is a regression estimate of how many items a child answered correctly in the multiplication section. X is the number of items the child answered correctly on the remaining subtest.

The regression equation was applied to the 70 children who had not taken the multiplication section, to estimate what their scores would have been on those items. These children were each given a "corrected subtest score," which was calculated by adding the estimated number of items right in the multiplication section to the actual number of items right on the rest of the subtest.

Adjustment of Student Achievement Scores

Literature promoting IGE suggests that student achievement will be raised by implementing the features of IGE (e.g., Klausmeier, Rossmiller, & Saily, 1975). Therefore, student achievement measures are an important part of the Phase I evaluation. One would expect schools high on implementing IGE features to have high achieving students. Examining the relations between IGE features and student achievement, however, can be misleading if differences due to other influences are not accounted for. Students from a school in a wealthy suburb, for example, would commonly score higher on an achievement test than students from a school in a poor urban neighborhood (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966). Some differences

in academic achievement between rich and poor children appear to persist even when their school programs are identical (Firkowska, Ostrowska, Soskolowska, Stein, Susser, & Wald, 1978). Achievement differences associated with community type cannot be ignored. Presumably, a major reason why students in rich and poor neighborhood schools would differ in achievement is that they are unequal at the outset in their command of the knowledge and skills required for successful school performance.

Schools are unevenly matched in terms of the student populations they serve. That fact is commonly recognized when schools are compared directly. Teachers and administrators from different schools are usually aware of the disparate backgrounds from which each other's students come. However, the need to control for differences in student background is less obvious when relationships between school characteristics and student achievement are being studied. The influence of a school characteristic or practice on achievement can be obscured if differences in student background are not controlled. An educational practice that has no effect on achievement can be correlated with achievement if most schools using the practice serve a population of educationally advantaged students. Conversely, an effective practice can appear ineffective if most schools using it serve a population of educationally disadvantaged students.

To control for differences in student achievement that result from differences in student populations, two measures were taken: the demographic and economic character of each school's community (demographic background), and student's academic ability.

Demographic background pertains to the school neighborhood as a whole. Consequently, it is not sensitive to educationally important differences that exist between the backgrounds of children who attend the same school. It simply represents the kind of neighborhood being served: suburban, inner city, small town, rural, etc. Therefore, demographic background was seen as a reasonable proxy for the school setting, the educational ambience of the community, and the prevailing economic circumstances of students' families. Student-specific measures of these differences would certainly have been more accurate, but they would also have been costly and intrusive on the privacy of students' families. A more complete discussion of the demographic background measure is given by Price, Janicki, Howard, Stewart, Buchanan, and Romberg (1978).

The Short Form Test of Academic Aptitude (SFTAA), a measure of scholastically useful skills and knowledge, was given to each child. This measure provided information about individual students' short-term prospects for success in school subjects. A high score on the SFTAA suggests that a student has acquired--through home, prior schooling, and elsewhere--skills and knowledge that make it easier to succeed in school. Since there are innumerable ways to get a low score on tests such as the SFTAA, the educational prospects of low scoring students are less clear. Nevertheless, a low score suggests that a student may lack skills and knowledge useful to school learning. The SFTAA, therefore, gives probabilistic information about scholastically useful skills and knowledge of students. Like any probabilistic information, it can be wrong in individual cases.

The academic ability scores of individual students can be used to predict how those students will fare on subject area achievement tests. That prediction, in fact, was our primary means of making adjustments for achievement test differences that could not be reasonably attributed to school program differences. However, prediction on the basis of information specific to the individual student reveals only part of the effect academic ability can have on achievement test performance. By aggregating the SFTAA scores of students in a school, a measure of the school's educational milieu is provided. This can augment the prediction provided by individual scores. In other words, one might predict different achievement from two students who obtained the same SFTAA score, if one attended a school in which most students scored low on the SFTAA and the other attended a school in which most scored high.

Student raw scores on achievement measures were carefully adjusted to provide a better comparison among schools. The student achievement variables used in subsequent analyses of the Phase I study were not simply raw scores, but residual scores derived from the differences between raw and predicted scores on achievement tests. Each student's scores on academic ability tests were used to predict his or her scores on the achievement tests. The student's actual scores on those measures were then compared with the predictions. The deviations of the actual scores from the predicted scores were the variables of interest. These difference scores, known technically as regression residuals, were then aggregated. The aggregation procedure and the reason for it are explained on page 17.

Guarding Against Faults in SFTAA Scores

The adjusted achievement scores, central to subsequent analyses, are susceptible to any inaccuracies that might exist in SFTAA scores. An unrepresentative score on the SFTAA would cause an inappropriate adjustment to be made on a student's achievement subtest scores. Therefore, special pains were taken to assure that this basis for adjustment had values which seemed reasonable.

One cause of unrepresentative SFTAA scores is mechanical errors in taking the test. Second grade pupils, who have little experience in taking paper-and-pencil tests, are likely to make such errors.

Faulty test administration is another possible source of difficulty. If a teacher's explanations or directions are unclear, student scores can be adversely affected. Some questions on the SFTAA are read to the students by the teacher, so some teacher-caused problems are inevitable. Errors in timing the tests could distort the scores considerably. Finally, a student might not achieve a representative score because of a bad day, home troubles, illness, test anxiety, discomfort about the testing room, or distractions in a group setting.

Winsorization. A procedure known as "winsorization" was used to help avoid problems that can beset a one-time measure of academic ability. Generally, winsorization involves replacing extreme scores with scores that are still on the outskirts of a distribution, but which are less extreme (see Dixon & Tukey, 1968). In effect, winsorization pulls improbably high or low scores back to a believable range. Unrepresentative low scores on the SFTAA occur much more frequently

than unrepresentative high scores, so only the lower tail of the distribution was winsorized for this study.

Thus far, we have described winsorization as a technique used only to adjust extreme values in a *univariate distribution*--that is, values of a particular variable that are so extreme as to be implausible. In addition, winsorization can be applied to outliers from a *bivariate distribution*. Implausible pair-wise *combinations of values* are singled out for winsorization in a bivariate distribution.

Winsorization with respect to bivariate distribution was the approach used in the present study. The two variables, whose joint distribution was adjusted for outliers, are raw scores from the two halves of the SFTAA: language ability and nonlanguage ability. These subtests are correlated, but it is common for individuals to have a higher normative standing on one subtest than on the other. It is presently fashionable in educational practice to interpret a large discrepancy between such scores as symptomatic of problems that can interfere with school learning. However, a test that is susceptible to misrepresenting a child's underlying competence is a poor basis for making inferences about such problems. Consequently, we make no psychoeducational interpretation of discrepancies between children's language ability and nonlanguage ability scores.

Some children will have authentically large differences between scores on two tests, yet for a particular population of children there are bounds within which discrepancies must stay to be credible. Discrepancies beyond those bounds can be caused by measurement procedures

that have gone awry. Alternatively, they can accurately describe a child who is so atypical that the evaluator should question whether it is taxonomically correct to include the child in the same population as the other children. Children who are very atypical should be removed from an analysis to avoid a distorted picture of the population. If there is insufficient evidence to justify such removal, then winsorization is appropriate. Winsorization does not eliminate large discrepancies, but it reduces them to a plausible level.

As mentioned previously, unrepresentative low scores were more likely to occur than unrepresentative high scores. Therefore, when the discrepancy between language ability and nonlanguage ability was implausibly large, the lower of the two scores was winsorized. The higher score was not. This asymmetrical approach tends to reduce the bivariate normality of a distribution by imparting a slight megaphone shape to the fringes of the bivariate distribution. However, since the effect on the shape of the distribution is not great, we decided to use the approach so our data would be more trustworthy. Scatterplots of the bivariate distribution before and after winsorization are shown in Figures 1 and 2.

Based on the bivariate distribution, 95% confidence bands for individual scores were calculated for language ability and nonlanguage ability, respectively. To make these calculations, each variable had to be regressed on the other. Scores falling too far below one of the regression lines were adjusted back to the lower bound of the confidence interval about that regression line. The formula for a 95% confidence

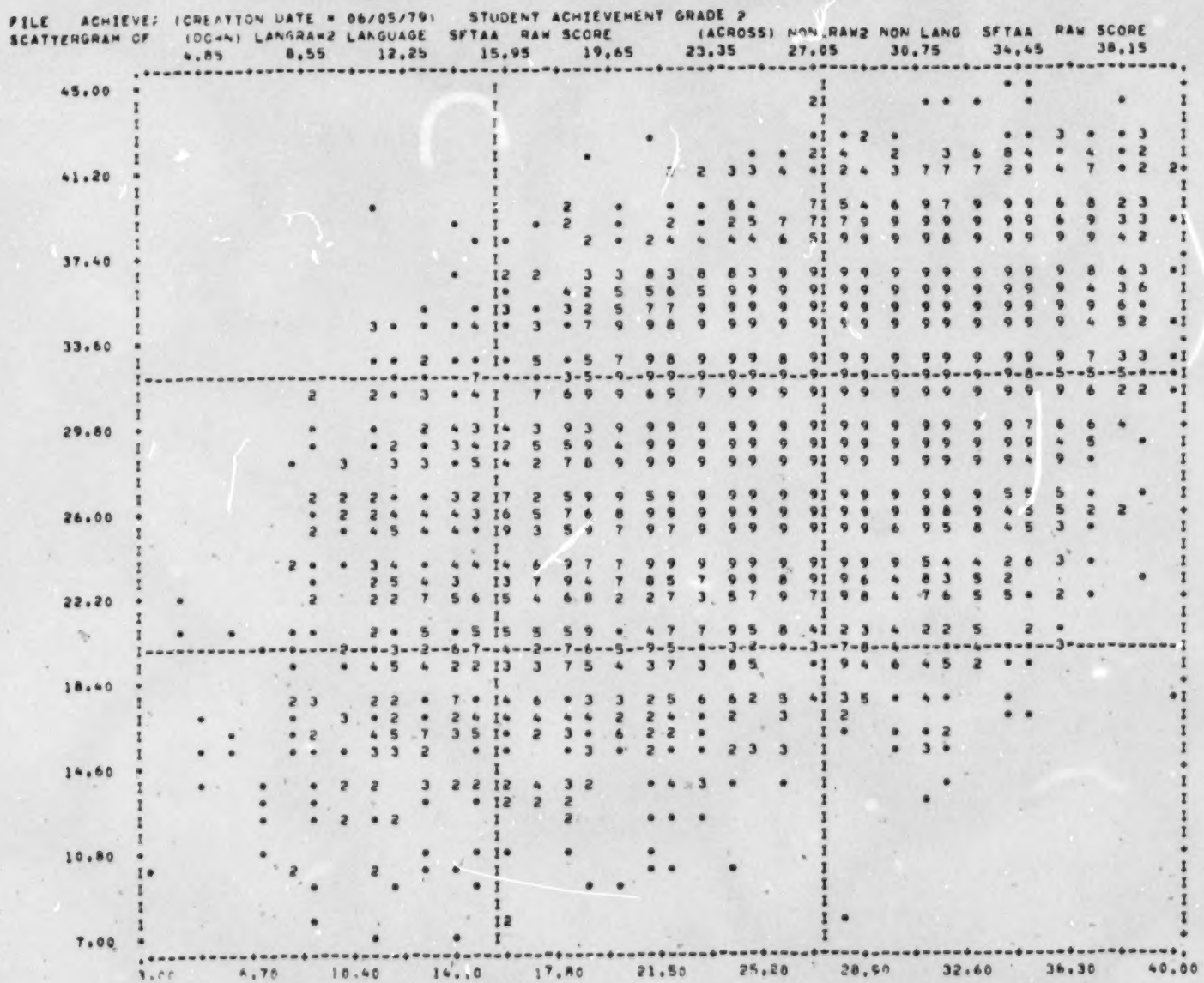

Figure 1. SFTAA bivariate distribution before winsorization.

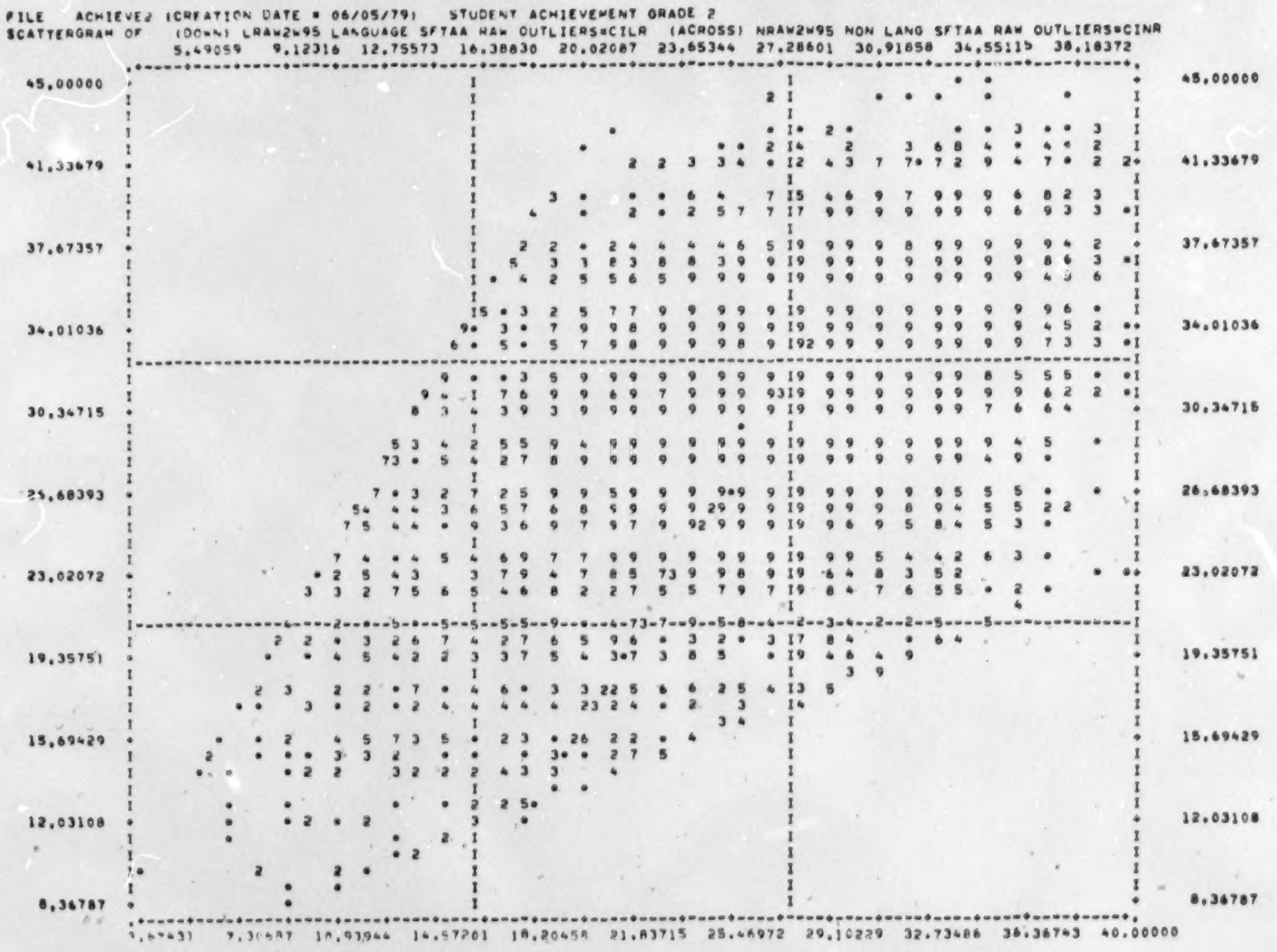

Figure 2. SFTAA bivariate distribution after winsorization.

to be regressed on the other. Scores falling too far below one of the regression lines were adjusted back to the lower bound of the confidence interval about that regression line. The formula for a 95% confidence band about a bivariate regression is given by Draper and Smith (1966, p. 24) as:

$$\text{lower .95 } CI = \hat{Y}_k - t(v,\ 0.975) \left\{ 1 + \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right\}^{1/2} S$$

We used the following equation, algebraically equivalent to that of Draper and Smith, to calculate the lower 95% confidence band for each variable:

$$\text{lower .95 } CI = \hat{Y}_k - 1.960\ s_y \sqrt{1 + \frac{1}{n} + \frac{\sum (X_k - \bar{X})^2}{n\ s_x^2}}$$

$\hat{Y}_k$ is the point estimation of Y for individuals having an X value of X_k; v is degrees of freedom for students' t; n is number of cases; $\bar{X}$ is the mean of X values; $\sum (X_i - \bar{X})^2$ is the sum of squared deviations; s_x is the standard deviation of X; and s_y is the standard deviation of Y. For each value on one variable there is a corresponding value on the other variable that defines the lower bound of its confidence interval.

In practice, students with low scores on both language and non-language ability, or with two high scores, did not have their scores adjusted. Only students with relatively good scores on one subtest and very poor scores on the other received adjusted scores. By adjusting only as far as the confidence band, the distortion of erroneous outliers was minimized. Out of 5,352 students, only the extremely discrepant scores of 242 students were affected by the winsorization procedure.

Achievement Scores as Regression Residuals

The SFTAA language ability and nonlanguage ability measures were then used in multiple regression equations to predict the CTBS achievement subtest scores of each student. Scores of all students who took an achievement subtest (*N* approximately 1,384) were used in the regression analysis. Each CTBS subtest (in raw score form) was regressed on SFTAA language ability and nonlanguage ability (also in raw score form) and on sex.[1] The unstandardized coefficients of these multiple regressions are shown in Table 2.

Each student's predicted score on each subtest was estimated from the regression equation. The predicted score was then subtracted from the student's actual raw score on the subtest, thereby creating a residual score for each subtest. This residual score reflects the student's achievement after being statistically adjusted to control for the effects of academic ability and sex. The rationale for doing this was given earlier.

---

[1]Performance on the SFTAA can be scaled in several ways. We used raw score scaling in preference to IQ scaling. When an ability test such as the SFTAA is being used to predict school achievement, its IQ scale form can yield unintended results--unless age is also included as a predictor.

Table 2

Grade 2 Multiple Regression Coefficients

CTBS achievement subtests	SFTAA Language	SFTAA Nonlanguage	Sex[a]	Intercept
Reading vocabulary	.405	.367	-1.842	1.771
Reading comprehension of sentences	.278	.261	-2.289	.240
Reading comprehension of passages	.243	.279	-1.810	-2.600
Mathematics computation	.140	.330	.292	.552
Mathematics concepts and applications	.226	.351	.335	- .905

[a]Boys = 1; girls = 0.

Aggregation of Regression Residuals

The residual scores of individual students on CTBS subtests provided the basis for an aggregated measure representing each school. The scores of students in each school were averaged to form a school mean for each subtest. In other words, the adjusted scores (regression residuals) of students in an I&R unit were averaged to provide a representative measure of (ability-adjusted) student achievement in that I&R unit. Figures 3 through 7 show how the means of 156 I&R units were distributed on five (regression-adjusted) achievement subtests. The subtests are, in sequence, reading vocabulary, reading comprehension of sentences, reading comprehension of passages, mathematics computation, and mathematics concepts and applications.

After aggregation, I&R unit means on residual scores from achievement subtests were combined to form an overall reading achievement score and an overall math achievement score. Recall that overall scores could not have been formed for individual students because, with a few exceptions, individuals took only some of the subtests. The overall reading and math achievement scores--formed by simple addition of I&R unit averages of ability-adjusted subtest scores--are the primary achievement measures used in later analyses. Figure 8 presents a histogram showing the distribution of I&R unit means on the overall adjusted reading achievement variable. Figure 9 presents an analogous histogram for math achievement.

Table 3 provides descriptive statistics about the univariate distributions of all ability and achievement measures, using the individual pupil as the unit of analysis. Table 4 also presents descriptive statistics, using the I&R unit as the unit of analysis.

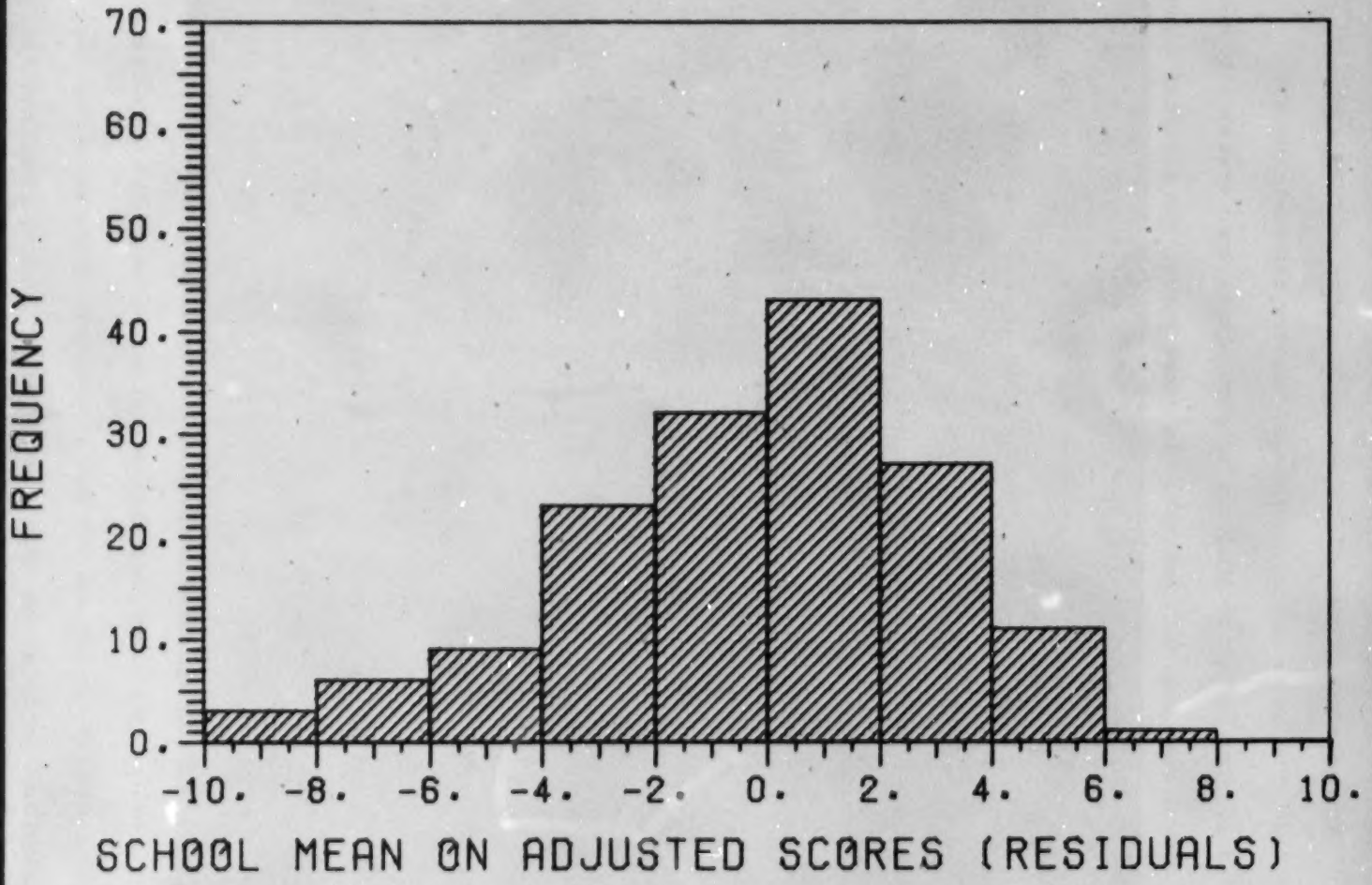

Figure 3. Reading vocabulary (CTBS) grade 2.

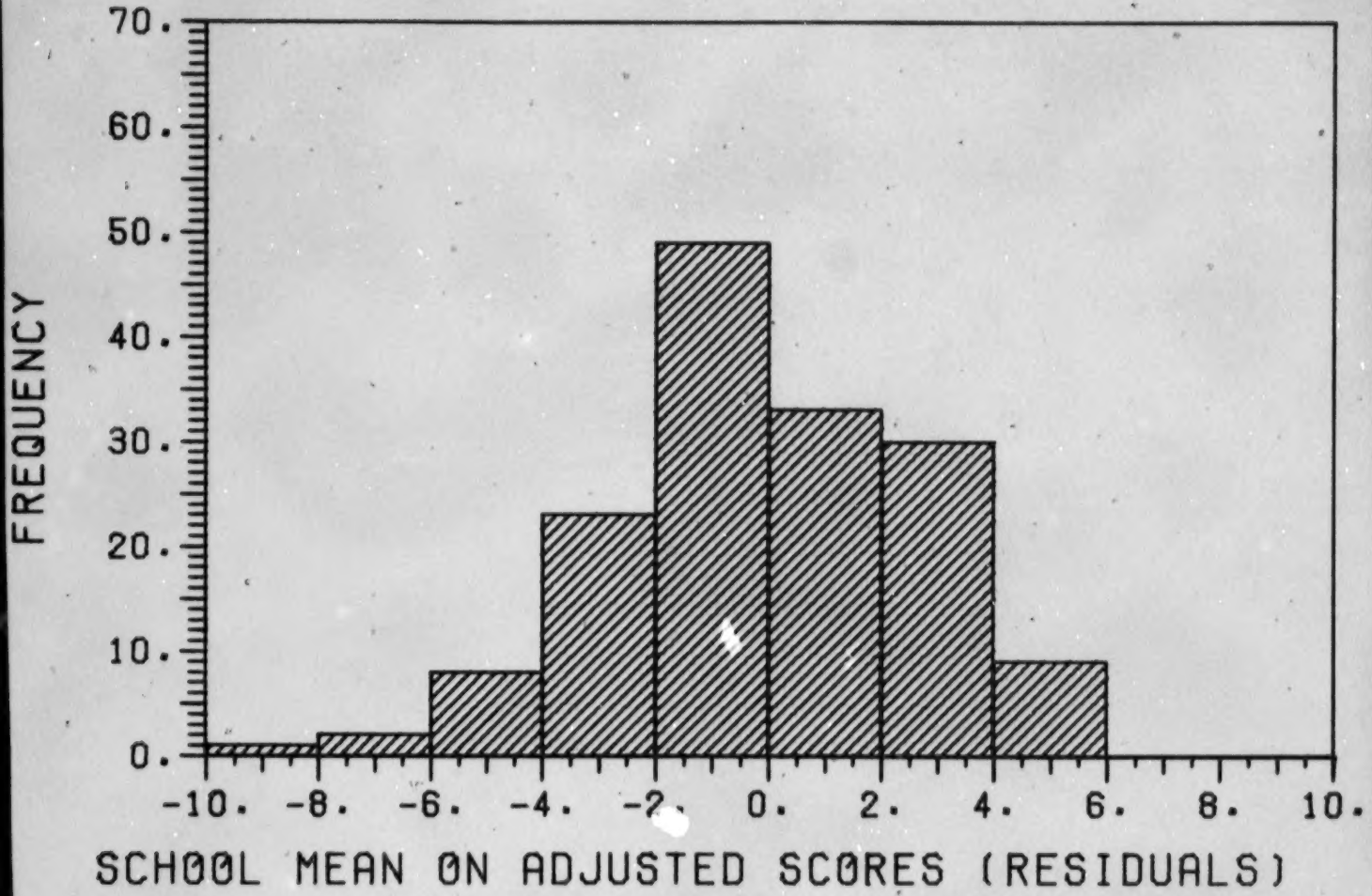

Figure 4. Reading comprehension of sentences (CTBS) grade 2.

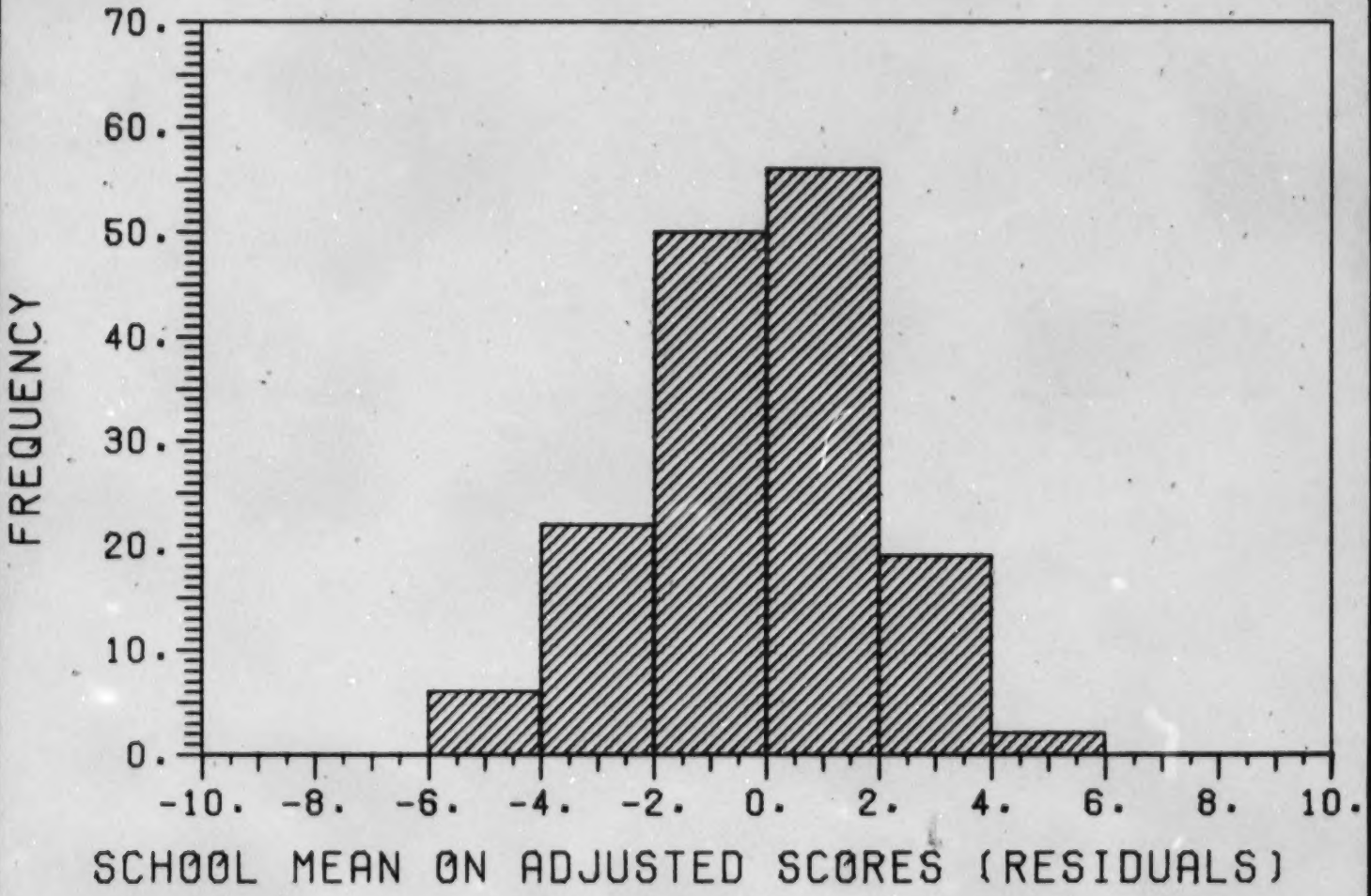

Figure 5. Reading comprehension of passages (CTBS) grade 00.

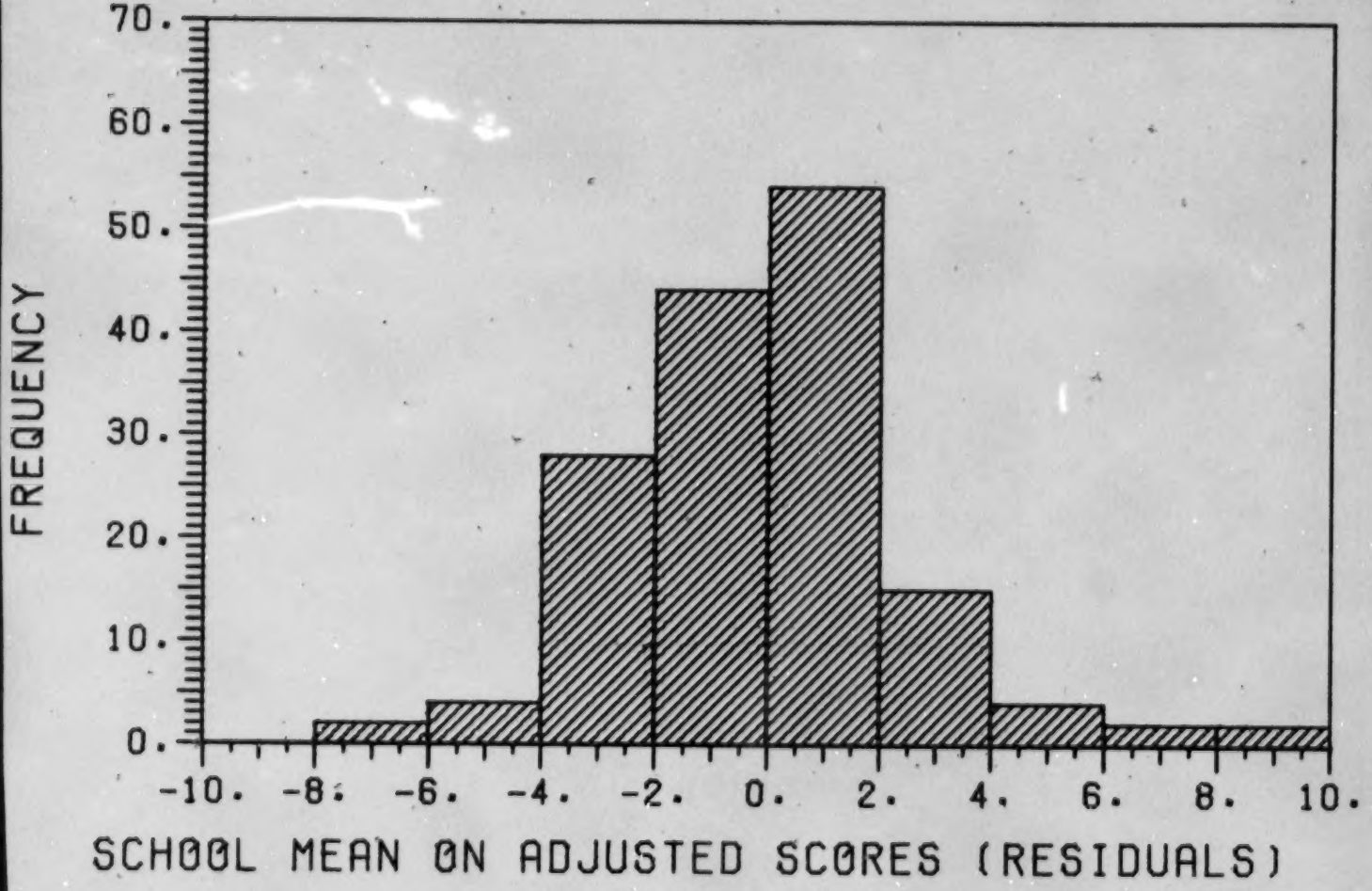

Figure 6. Mathematics computation (CTBS) grade 2.

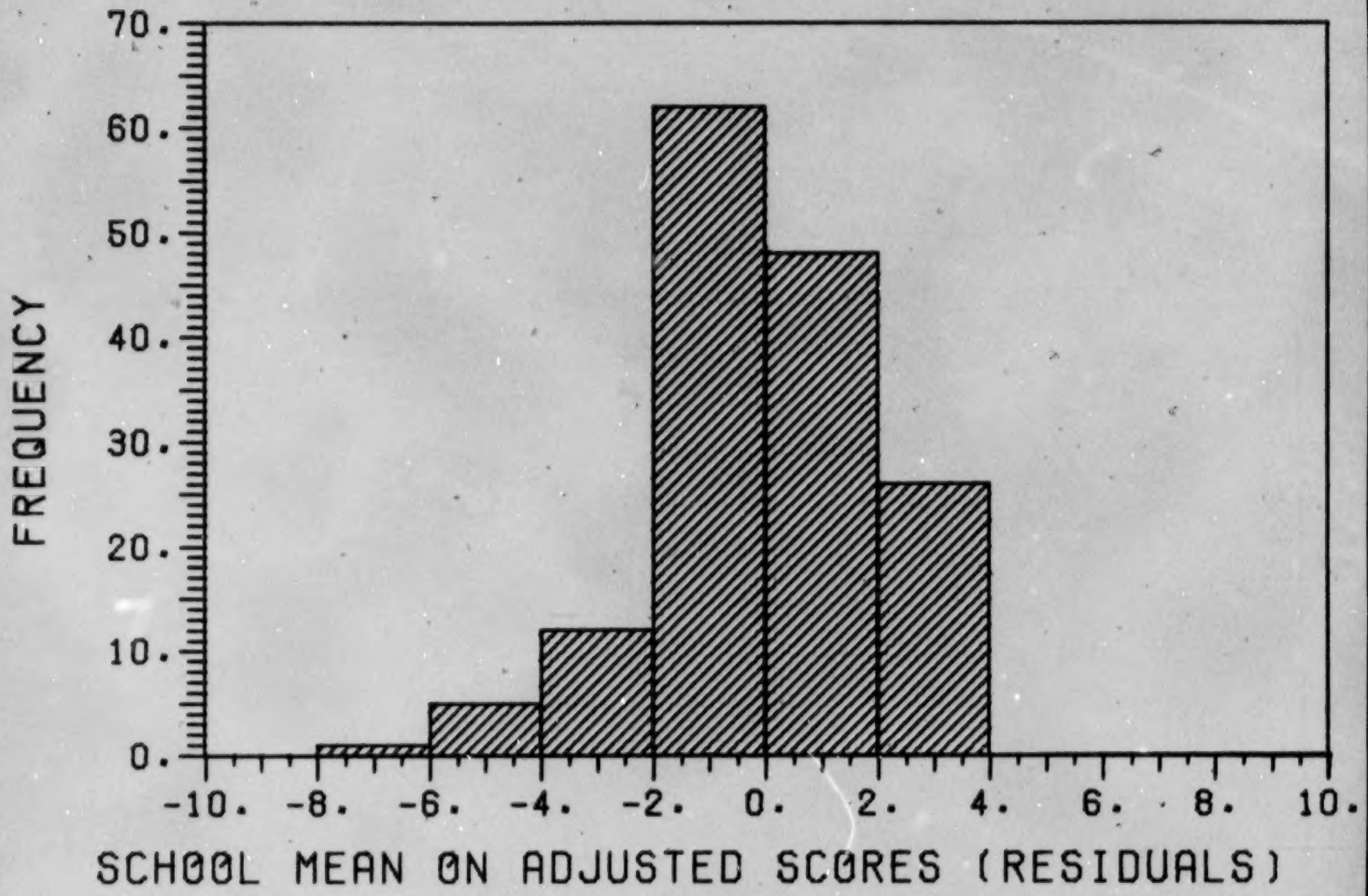

Figure 7. Mathematics concepts and applications (CTBS) grade 2.

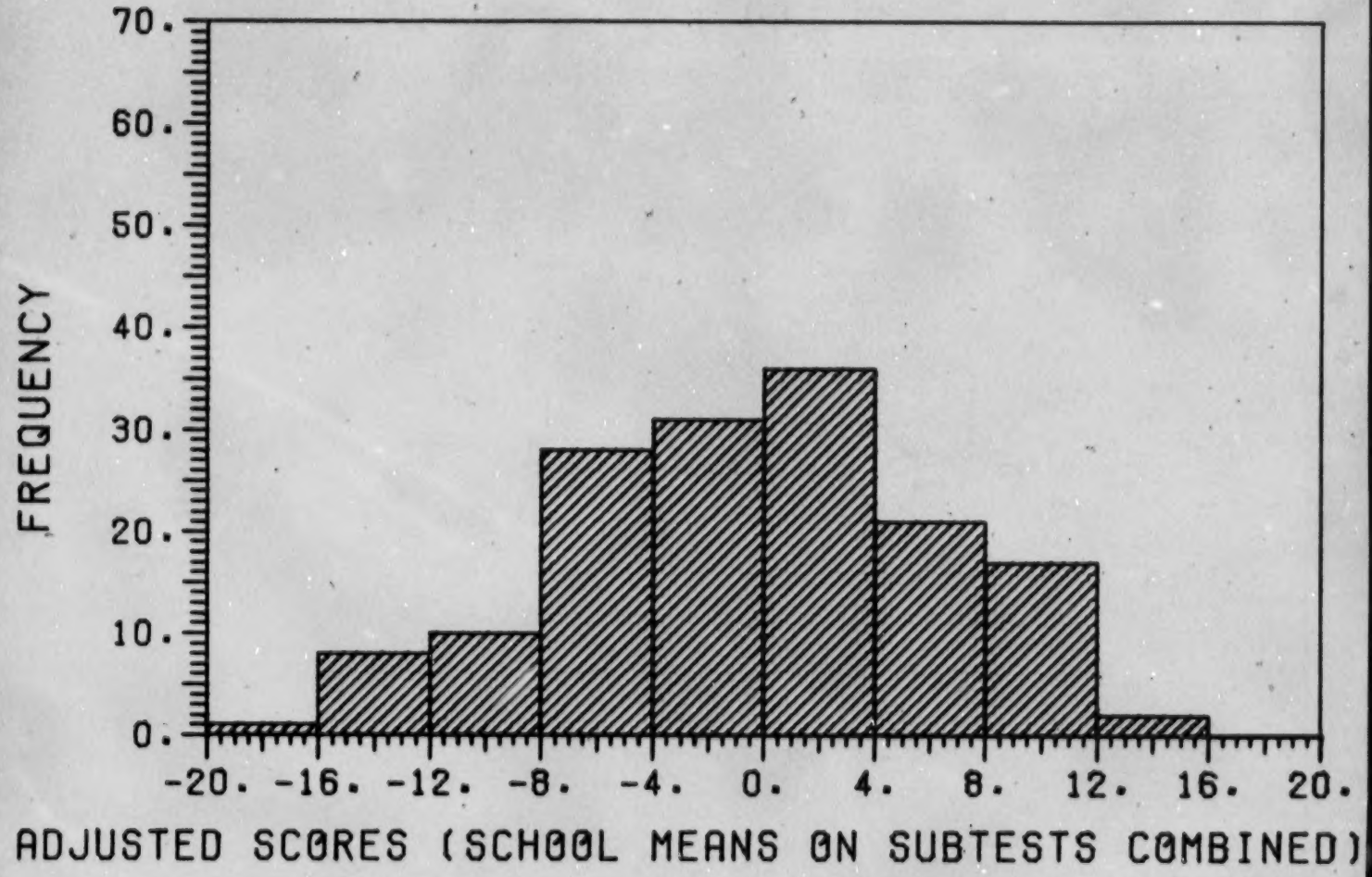

Figure 8. Adjusted reading achievement grade 2.

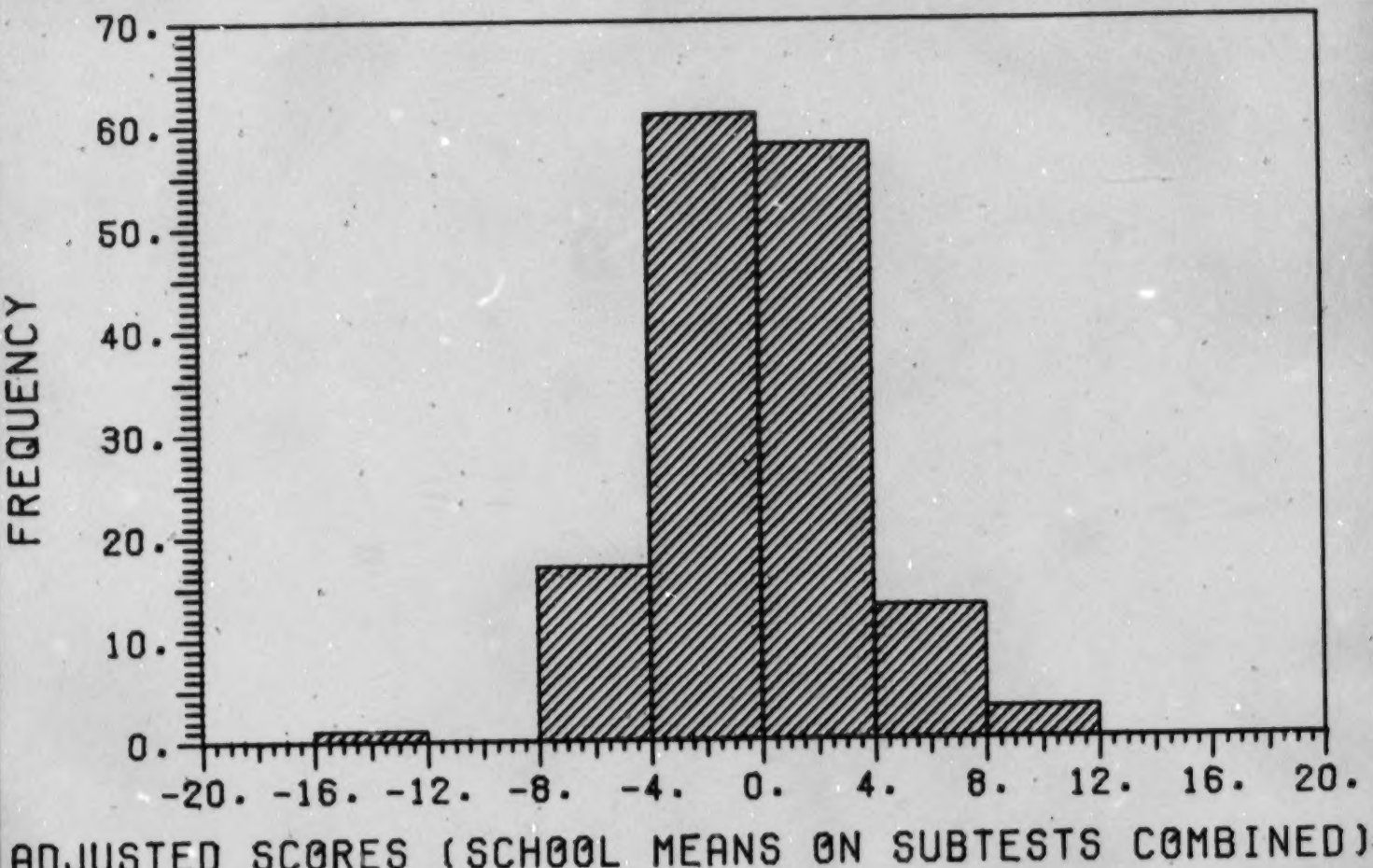

Figure 9. Adjusted mathematics achievement grade 2.

Table 3

Grade 2 Student Ability and Achievement

Variable	Valid *N*	Mean	Standard deviation	Minimum	Maximum
Short Form Test of Academic Aptitude (SFTAA)					
Language ability (raw score)	5,346	29.688	6.517		
Language ability (adjusted by winsorization)	5,349	29.724	6.436	8.368	45.000
Nonlanguage ability (raw score)	5,312	26.732	6.678		
Nonlanguage ability (adjusted by winsorization)	5,349	26.747	6.583	8.674	40.000
Comprehensive Test of Basic Skills (CTBS)					
Reading vocabulary	1,391	22.694	7.667	4.000	33.000
Reading vocabulary (adjusted)	1,380	- .133	6.210	-22.347	15.158
Reading comprehension of sentences	1,391	14.352	6.939	1.000	23.000
Reading comprehension of sentences (adjusted)	1,380	- .122	6.130	18.135	12.758
Reading comprehension of passages	1,367	11.191	5.539	.000	18.000
Reading comprehension of passages (adjusted)	1,352	- .034	4.616	-13.946	11.643
Mathematics computation	1,376	13.688	5.643	1.000	28.000
Mathematics computation (adjusted)	1,362	- .044	4.924	-12.156	17.047
Mathematics concepts and applications	1,402	15.381	5.072	2.000	25.000
Mathematics concepts and applications (adjusted)	1,390	- .025	3.857	-12.432	13.065

Table 4

Grade 2 Student Ability and Achievement

Descriptive Statistics for School Aggregates

Variable	Valid *N*	Mean	Standard deviation	Minimum	Maximum
Short Form Test of Academic Aptitude (SFTAA)					
Language ability (adjusted by winsorization)	156	29.794	2.929	19.708	36.944
Nonlanguage ability (adjusted by winsorization)	156	26.814	2.865	16.253	32.000
Comprehensive Test of Basic Skills (CTBS)					
Reading vocabulary	155	22.683	4.071	8.167	30.500
Reading vocabulary (adjusted)	155	- .168	3.123	-9.521	6.521
Reading comprehension of sentences	155	14.327	3.132	6.000	21.308
Reading comprehension of sentences (adjusted)	155	- .152	2.669	-8.081	5.547
Reading comprehension of passages	155	11.107	2.814	3.429	16.333
Reading comprehension of Passages	155	- .121	2.065	-5.244	5.645
Mathematics computation	155	13.619	3.036	5.625	23.500
Mathematics computation (adjusted)	155	- .121	2.545	-7.355	10.286
Mathematics concepts and applications	154	15.375	2.715	8.571	21.100
Mathematics concepts and applications (adjusted)	154	- .041	1.870	-6.323	3.647
Reading total score	154	48.114	8.985	21.000	65.538
Reading total score (adjusted)	154	- .454	6.687	-18.909	12.402
Mathematics total score	153	29.007	4.869	16.054	43.856
Mathematics total score (adjusted)	153	- .260	3.470	-19.857	7.627

References

Bialer, I. Conceptualization of success and failure in mentally retarded and normal children. Journal of Personality, 1961, 29, 303-320.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. P., & York, R. L. Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Dixon, W. J., & Tukey, J. W. Approximate behavior of the distribution of winsorized t (trimming/winsorization 2). Technometrics, 1968, 10, 83-98.

Draper, N. R., & Smith, H. Applied regression analysis. New York: John Wiley & Sons, 1966.

Firkowska, A., Ostrowska, A., Soskolowska, M., Stein, Z., Susser, M., & Wald, I. Cognitive development and social policy. Science, 1978, 200, 1357-1362.

Klausmeier, H. J., Rossmiller, R. A., Saily, M. (Eds.). Individually guided elementary education: Concepts and practices. New York: Academic Press, 1977.

Klopp, P. M., Buchanan, A. E., Stewart, D. M., & Romberg, T. A. Conduct of the study: Phase I of the IGE evaluation project. (Working Paper 258). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Price, G. G. Sampling for Phase I of the IGE evaluation (Working Paper 233). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1977.

Price, G. G., Janicki, T. C., Howard, J. A., Stewart, D. M., Buchanan, A. E., & Romberg, T. A. Procedures used in Phase I of the IGE evaluation to scale the variable School's Demographic Setting (Technical Report 482). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Romberg, T. A. IGE evaluation: Perspectives and a plan (Working Paper 183). Madison: Wisconsin Research and Development Center for Cognitive Learning, 1976.

Stewart, D. M. Instrumentation for Phase I of the IGE evaluation (Unpublished Project Paper 77-1). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1977.

Center Planning and Policy Committee

Richard A. Rossmiller
Wayne Otto
Center Co-Directors

Dale D. Johnson
Area Chairperson
Studies in Language:
Reading and Communication

Marvin J. Fruth
Area Chairperson
Studies in Implementation
of Individualized Schooling

Penelope L. Peterson
Area Chairperson
Studies of Instructional Programming
for the Individual Student

James M. Lipham
Area Chairperson
Studies of Administration and
Organization for Instruction

Thomas A. Romberg
Area Chairperson
Studies in Mathematics and Evaluation
of Practices in Individualized Schooling

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Associate Professor
Curriculum and Instruction

W. Patrick Dickson
Assistant Professor
Child and Family Studies

Lloyd E. Frohreich
Associate Professor
Educational Administration

Marvin J. Fruth
Professor
Educational Administration

Dale D. Johnson
Professor
Curriculum and Instruction

Herbert J. Klausmeier
V.A.C. Henmon Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

James M. Lipham
Professor
Educational Administration

Dominic W. Massaro
Professor
Psychology

Donald M. McIsaac
Professor
Educational Administration

Wayne Otto
Professor
Curriculum and Instruction

Penelope L. Peterson
Assistant Professor
Educational Psychology

Thomas S. Popkewitz
Professor
Curriculum and Instruction

Gary G. Price
Assistant Professor
Curriculum and Instruction

W. Charles Read
Professor
English and Linguistics

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Peter A. Schreiber
Associate Professor
English and Linguistics

B. Robert Tabachnick
Professor
Curriculum and Instruction

Gary G. Wehlage
Professor
Curriculum and Instruction

Louise Cherry Wilkinson
Associate Professor
Educational Psychology

Steven R. Yussen
Professor
Educational Psychology

11/79